



Zhang, Q., & Damian, M. (2019). Syllables constitute proximate units for Mandarin speakers: Electrophysiological evidence from a masked priming task. *Psychophysiology*, 56(4), [e13317].
<https://doi.org/10.1111/psyp.13317>

Peer reviewed version

Link to published version (if available):
[10.1111/psyp.13317](https://doi.org/10.1111/psyp.13317)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at <https://onlinelibrary.wiley.com/doi/full/10.1111/psyp.13317>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Syllables constitute proximate units for Mandarin speakers:

Electrophysiological evidence from a masked priming task

Qingfang Zhang¹ Markus F. Damian²

¹Department of Psychology, Renmin University of China, Beijing, 100872, China

²School of Psychological Science, University of Bristol, UK

Address correspondence to:

Qingfang Zhang

Department of Psychology, Renmin University of China

59 Zhongguancun Street, Haidian District, Beijing 100872, China

Email: qingfang.zhang@ruc.edu.cn

Abstract

Languages may differ regarding the primary mental unit of phonological encoding spoken production, with models of speakers of Indo-European languages generally assuming a central role for phonemes, but spoken Chinese production potentially attributing a more prominent role to syllables. In the present study, native Mandarin Chinese speakers named objects which were preceded by briefly presented and masked prime words which were form related and either matched or mismatched concerning their syllabic structure, or were unrelated. Behavioral results showed a previously reported interaction between prime and target syllable type. Concurrently recorded electroencephalography (EEG) also exhibited this interaction, and further revealed that syllable overlap modulated ERPs mainly in the time window of 300-400 ms after picture onset. By contrast, phonemic overlap modulated ERPs from 500 ms to 600 ms. This pattern might suggest that speakers retrieved syllables before phonemes, and strengthens the claim that for Chinese individuals, syllables constitute primary functional representations (“proximate units”).

Key words: spoken production, syllable priming effect, phonemic overlap effect, proximate units hypothesis

1. Introduction

Speaking involves conceptual preparation, lexical access, word-form encoding and articulatory processes (Caramazza, 1997; Dell, 1986, 1988; Humphreys, Riddoch, & Quinlan, 1988; Levelt, 1989; Levelt, Roelofs, & Meyers, 1999). In the model of language production advocated by Levelt and colleagues, word-form encoding refers to the retrieval of morphemes in the mental lexicon, and is further divided into segmental (i.e., phoneme-based) and metrical spell-out levels (Levelt et al., 1999; Levelt & Wheeldon, 1994; Roelofs, 1997a, b). Studies have investigated this stage with the aim of identifying functional units that underlie phonological encoding. In past few decades, the role played by sub-lexical units in this process such as the phoneme, syllable or mora has received much attention (J.-Y. Chen, Lin, & Ferrand, 2003; O'Seaghdha, J.-Y. Chen, & T.-M. Chen, 2010; Schiller, 1998, 1999, 2000; Verdonschot, Kiyama, Tamaoka et al., 2011). Most frameworks of language production assume the existence of “phonemes”, i.e., abstract mental codes corresponding to individual speech sounds which form the inventory of a given language, as well as syllable-sized production units which play their role either as abstract frames (e.g., Costa & Sebastian-Galles, 1998) or as articulatory gestural scores (Levelt & Wheeldon, 1994; Levelt et al., 1999).

Evidence for phonemes as central units in spoken production comes from speech error analyses (Dell, 1986; Shattuck-Hufnagel, 1979) which suggest that most phonological errors involve the insertion, deletion, substitution and exchange of phonemes. By contrast, errors involving whole syllables are relatively rare. Some converging evidence which underscores the importance of phonemes comes from error-free spoken production. For instance, in a form preparation (“implicit priming”) task, the central finding with speakers of various Indo-European languages (e.g., Dutch: Meyer, 1991; French: Alario, Perre, Castel, & Ziegler, 2007; English: Damian & Bowers, 2003) is that

spoken responses are faster when response words overlap in terms of shared phonemes, compared to when they do not. Studies using a “phoneme repetition” task in which speakers named colored objects also found faster naming latencies when adjective and noun shared phonemes (“green goat”; “red rug”) than when not (“red goat”; “green rug”; Damian & Dumay, 2007, 2009). Findings such as these provide evidence for phonemes as important phonological units of spoken production.

Relevant evidence concerning the status of the syllable as a functional unit of speaking comes from studies which have investigated potential syllable frequency effects. Levelt and Wheeldon (1994) were the first to explore the logic that, if syllables are stored units of production, then syllable frequency effects should be detectable in the latencies of spoken responses. In their pioneering studies, effects of syllable frequency indeed emerged for Dutch disyllabic response words, although mainly arising from the frequency of the second syllable. More recently, syllable frequency effects have also been shown in the production of Dutch (Cholin, Levelt, & Schiller, 2006) and Spanish (Carreiras & Perea, 2004) pseudowords, but effects mainly arose from the frequency of the initial syllable. Most recently, Cholin, Dell and Levelt (2011) demonstrated syllable frequency effects in the naming of English mono- and disyllabic pseudowords; for disyllabic responses, the frequency of both syllables appeared to affect naming latencies.

Other than via explorations of syllable frequency effects, relevant evidence mainly stems from “masked priming” tasks, and here the evidence is somewhat more mixed. In these studies, to-be-named target pictures or words are preceded by very briefly presented and masked prime words or nonwords, such that the primes are difficult to perceive. Form overlap between prime and target is manipulated, and the aim is to see which kind, and degree, of overlap affects target naming latencies. One such overlap type is the match between prime and target in terms of syllabic

structure. Ferrand and colleagues (Ferrand, Segui, & Grainger, 1996; Ferrand, Segui, & Humphreys, 1997) reported a series of experiments in which speakers named pictures, words, or nonwords, and masked prime matched or mismatched regarding their initial syllable structure. For instance, in Ferrand et al. (1996), French participants were asked to name words or pictures with target names starting with a CV syllable, e.g., *pa.lace* (*palace* in English, a dot indicates a syllable boundary, similarly hereinafter), or with a CVC syllable, e.g., *pal.mier* (*palm* in English). Targets were preceded by a CV prime, e.g., *pa* or a CVC prime, e.g., *pal*, which were presented very briefly (29 ms) to prevent conscious perception. The central finding was an interaction between target and prime type: CV targets (e.g., *pa.lace*) were named faster when preceded by CV primes (*pa*) than by CVC primes (*pal*), and CVC targets (e.g., *pal.mier*) were named faster when preceded by CVC primes (*pal*) than by CV primes (*pa*). Similar effects of syllable structure with French speakers were subsequently reported by Chetail and Mathey (2009). Ferrand et al. (1997) reported equivalent syllabic effects with English speakers and materials. A masked syllable priming effect of this type supports the notion that the syllable constitutes an essential unit of word-form encoding in speech production. However, other studies did not find syllable-based masked priming (e.g. Schiller, 1998 in Dutch; Schiller, 1999, 2000 in English). Brand, Rey and Peereman (2003) failed to replicate Ferrand et al.'s (1996) original findings even when using identical stimuli, experimental design, and the same target language of French. Schiller, Costa and Colomé (2002) found no or only weak syllable priming with French and Spanish speakers. Instead, these studies generally found that the magnitude of priming was determined exclusively by the amount of phonemic overlap between primes and targets, whereas match or mismatch regarding syllable structure was not relevant. Overall, in masked priming tasks which involve spoken production, priming arises due to prime-target form overlap in terms of phonemes,

but it appears difficult to reliably obtain syllabic effects. Broadly, this suggests that for speakers of the Indo-European languages targeted in these studies (Dutch, English, French, Spanish, etc.), phonemes constitute primary mental units of phonological encoding, whereas syllables might take on a more subordinate role.

The claim that phonemes are primary phonological encoding units in Indo-European languages is particularly plausible because the alphabetic orthographies of these languages explicitly code for phonemes via letters or letter combinations. What about non-alphabetic languages such as Chinese? Chinese employs a logographic writing script in which orthographic representations code not for phonemes, but instead characters map onto syllables.¹ It is hence plausible that for Chinese speakers, syllables may have particular prominence as units of phonological encoding, whereas phonemes are relatively less important. A second reason to expect a particularly important role of syllables in Chinese speakers is that compared to most Indo-European languages, spoken Mandarin Chinese has a very straightforward syllable structure with relatively few syllable types (about 400 not counting tone and 1200 counting tone), and no resyllabification (i.e., syllable boundaries are never adjusted based on lexical context, as is the case for many other languages with more complex syllable structures).

Recently emerging evidence offers some support for the possibility that for Chinese speakers, syllables rather than phonemes constitute primary units of phonological encoding. This possibility is for instance reflected in speech errors in which the target word 清浊度 (/qing1zhuo2du4/, clarity) was mistakenly pronounced as /qing1du2du4/, an anticipation of the entire third syllable /du/ (J.-Y.

¹ It is acknowledged that many characters contain a phonetic component which provides probabilistic cues to the character's pronunciation; nonetheless, Chinese orthography does not explicitly represent phonemes.

Chen, 2000). In the “implicit priming” task, word-initial phonemic overlap which generally causes priming in Indo-European languages (e.g., Meyer, 1991) results in a null finding with Chinese speakers, with priming only found when word-initial syllables overlap (J.-Y. Chen, T.-M. Chen & Dell, 2002). In a masked priming task of the type summarized in the previous passage, J.-Y. Chen et al. (2003) investigated the role of briefly presented prime words which matched or mismatched regarding their syllable structure with the spoken response words. In their third experiment, they found that CV targets were named faster when preceded by CV primes compared to CVG (G represents glide sound) primes, whereas the opposite pattern was obtained for CVG targets. This finding supports the notion that the syllable is a functional unit in Mandarin spoken word production. Similar results were subsequently reported by You, Zhang, and Verdonschot (2012) who also found a syllable priming effect with Mandarin speaker, in both word reading and picture naming. Finally and most recently, J.-Y. Chen, O’Seaghdha and T.-M. Chen (2016) reported results from masked priming experiments which included both syllabically and phonemically related primes. They found syllabic priming, but little evidence of phonemic priming; in fact, targets preceded by phonemically related primes were named slightly *slower* than targets preceded by unrelated primes. Overall, existing results from masked priming underscore the suggestion that syllables take on a primary role for Chinese speakers, but that phonemes are relatively of less importance.

If syllables have particular importance to Chinese individuals, do they mentally represent subsyllabic units such as phonemes or rhymes in speech production?² As already mentioned,

² Note that we restrict our scope to spoken production. For speech perception (rather than production) in Chinese, Luo et al (2006) found that with an odd-ball paradigm, a consonant contrast produced a stronger preattentive response in the left hemisphere than in the right hemisphere. This suggests that Chinese individuals represent phonemes when perceiving speech (see also Meng, Sai, Wang et al., 2005).

“implicit priming” studies generally show that priming is restricted to syllabic overlap (e.g., J.-Y. Chen et al., 2002). However, a few recent studies using other tasks have reported sub-syllabic priming effects in Chinese word production. Using the picture-word interference task and speakers of Cantonese, Wong and H.-C. Chen (2008) found significant facilitation effects when the target and the distractor word shared a syllable, or when they shared the syllable plus tone. Critically, these authors also observed significant facilitation when only the rhyme was shared (Experiment 5 using auditory distractors) or when rhyme plus tone were shared (Experiment 2 using visual distractors) (see Wong & H.-C. Chen, 2009, for related results, as well as Verdonschot, Lai, Feng, Tamaoka, & Schiller, 2015, for results with Chinese-English bilinguals). In a study reported by Qu, Damian and Kazanina (2012), Mandarin native speakers named colored line drawings of objects using adjective-noun phrases, and color and object name either shared the initial phoneme, or not. Behavioral data showed no effect of phoneme overlap, but concurrently recorded electroencephalography (EEG) revealed that phoneme repetition modulated event-related potentials (ERPs) beginning at 200 ms after picture onset (see also Yu, Mo & Mo, 2014, for similar findings from a task in which two pictures were successively named). Qu et al. (2012) and Yu et al. (2014) proposed that an earlier component of the ERP difference reflects a facilitatory effect of phoneme repetition, while a later ERP difference component reflects self-monitoring of phonemes. However, O'Seaghdha (2015; see also O'Seaghdha, J.-Y. Chen, & T.-M. Chen, 2013) challenged the interpretation of the phonemic ERP effect by suggesting that it might reflect the co-activation of the same onset for two words, rather than implying phonemes as a functional units of phonological encoding. Finally, a recent EEG study reported by Wang, Wong, Wang, and H.-C. Chen (2017) found evidence for syllables but not for phonemes: with a delayed picture naming task, syllable repetition elicited an ERP effect in the time

windows of 200-400 ms and 400-600 ms, whereas no significant ERP effects were found for phoneme repetition.

The array of findings summarized above undoubtedly paint a complex picture, but a plausible scenario is that speakers of all languages mentally represent phonemes and syllables, but attribute differential importance to each depending on the target language. O'Seaghdha and colleagues (O'Seaghdha et al., 2010; O'Seaghdha, 2015) have proposed a *proximate unit* principle according to which the first selectable phonological unit below the word can vary across languages. The studies summarized above suggest that proximate units in Indo-European languages are phonemes, whereas they are syllables in Mandarin Chinese. According to O'Seaghdha et al., word-form encoding in Mandarin Chinese consists of a number of steps schematized in Figure 1. Content and structure are separated in the model. Activation from the corresponding abstract word flows to phonological content and structure, for example, /hua4/ means the syllable /hua/ and its tone value. Phonological content is activated in a syllable while its syllabic frame is retrieved. Both syllabic content and frame are linked sequentially and tone is also specified at this point. Crucially, syllabic "chunks" are retrieved at an early stage of word form encoding, and because syllables constitute the proximate units for this target language, priming in experimental tasks is largely restricted to syllabic overlap. After the retrieval of syllables, the phonemes of the syllable are retrieved in parallel and linked to their position in the syllable frame. It is important to recognize the extent to which this model differs from those previously advocated for spoken production in Indo-European languages. For instance, in the model advocated by Levelt and colleagues (Levelt, 1989; Levelt et al., 1999), a word/morpheme directly activates a string of serially ordered phonemes (hence phonemes constitute "proximate units") and syllables are only accessed at a later stage which consists of articulatory gestural

encoding for articulation.

(Insert Figure 1 here)

The current study

The possibility that spoken production might differ in interesting ways across languages is clearly of high importance because it suggests that assumptions about how speakers plan and generate language in one language (or language family) cannot be uncritically extended to other languages (or language families). In the current project, we attempted to push the issue further, by conducting a masked priming task with Chinese speakers, and combining it with the measurement of EEG. We adapted the behavioral task from a previous study which had used masked priming to explore Chinese production (You et al., 2012, Experiment 3). An extensive literature exists which has explored effects of syllable match/mismatch (Cutler, Dahan, & van Donselaar, 1997; Cutler, Mehler, Norris & Segui, 1986; Mehler, Dommergues, Frauenfelder, & Segui, 1981). Most of these studies explored speech perception rather than production, but the logic is generally that the importance of syllabic structure is reflected in a crossover interaction between prime syllable type, and target syllable type (it is also possible to further collapse prime and target combinations into “syllable match” or “mismatch”, as e.g. in J.-Y. Chen et al., 2016). This also holds for the masked priming studies of spoken production summarized earlier, in which the syllable match or mismatch between a masked prime word, and an object name, is manipulated. Based on the earlier findings (J.-Y. Chen et al., 2003; You et al., 2012) we expected to find a syllabically based priming effect in response latencies (i.e., a crossover interaction between prime and target syllable type).

In a masked priming task, a potential phonemically based priming effect can be captured via

inclusion of an additional “neutral” prime condition. In most of the earlier studies on syllable priming such as J.-Y. Chen et al. (2003), no such unrelated condition was included. However, You et al. (2012) included an unrelated condition in which target stimuli were preceded by an unrelated Chinese character. In their Experiment 3 on which the current study is based, there was numerically little sign of phonemically based priming: latencies for targets with phonemically related (but syllable-mismatching) primes (i.e., CV targets preceded by CVN primes, or CVN targets preceded by CV primes; N represents nasal sound) were virtually identical to those in the neutral condition. In a more recent study, J.-Y. Chen, O’Seaghdha and T.-M. Chen (2016) re-paired primes and targets, separately for both their syllable and phoneme overlap conditions such that primes and targets were no longer phonologically related. As outlined above, they found that response latencies for the phoneme overlap condition were slightly slower than for the corresponding unrelated condition. These null, or perhaps weakly inhibitory, effects of phonemic overlap are in line with evidence from related tasks: as reviewed above, with Chinese speakers, phonemic overlap generally does not result in behavioral priming (e.g., J.-Y. Chen et al., 2002). In our study, we included a neutral condition to ensure comparability with the earlier experiment by You et al., but also because this allowed us to explore potential effects of phonemic overlap which might emerge in ERPs, which in Chinese are sometimes found in the absence of latency effects (Qu et al., 2012; Yu et al., 2014).

Combining the masked priming manipulation with electrophysiological measures additionally provides a fine-grained temporal estimate of information processing as speakers plan their response. We expected the critical interaction between prime and target type regarding their syllabic structure also to appear in ERP measures, such as the mean amplitude within a given time window. Specifically, the mean amplitude for CV targets were expected to be smaller when preceded by CV primes

compared to CVN primes, whereas the opposite pattern should be obtained for CVN targets.

Concerning the time course, based on the assumption that the syllable constitutes the “proximate unit” for spoken Mandarin Chinese, we predicted that syllable effects should emerge in ERPs in relatively “early” time windows following target presentation. Specifically, Indefrey and Levelt (2004) suggested a time window of 275-455 for the phase of phonological encoding in picture naming. This estimate was based on a meta-analysis of various relevant EEG-based findings; e.g., frequency effects (which are typically attributed to phonological encoding) are found within this time window (e.g., Laganaro, Morand & Schnider, 2009), as are so-called “cognate effects” in bilingual production (facilitation for words which are similar across the languages; Christoffels, Firk & Schiller, 2007). With regard to potential effects of phonemic overlap, comparison of the “phonemically related” to the “unrelated” condition would potentially allow us to isolate the effect and time point of phonemic encoding. Again based on the claim that in Chinese, syllables are “proximate units” whereas phonemes are accessed only later (see Figure 1), we anticipated that phonemic overlap should emerge in ERPs at a “late” point, relative to the emergence of syllabically based effects.

2. Experiment

2.1 Participants

Twenty-eight undergraduate students (10 males, average age = 22 years) participated in the experiment and were paid approximately \$12. They were recruited from Beijing Forestry University, China Agriculture University, and Beijing Science and Technology University. All were native Mandarin Chinese speakers with normal or corrected-to-normal vision.

2.2 Materials

The stimuli were identical to You et al.'s (2012) study. Forty black-on-white line drawings with disyllabic names were chosen as targets. Half of them had an initial CV syllable (CV targets), e.g., 鼻子 (/bi2.zi5/, *nose*), and half had an initial CVN syllable (CVN targets), e.g., 斑马 (/ban1.ma3/, *zebra*). CV and CVN target names were matched on agreement, on word frequency (Chinese Linguistic Data Consortium, 2003), on the frequency of the first character, the syllable frequency of the first character (Modern Chinese Frequency Dictionary, 1986), and on stroke number of the first character (see Table 1).

(Insert Table 1 here)

Eighty single Chinese characters were selected as primes, half consisting of CV syllables and half of CVN syllables (see Table 1 for properties). Each CV and CVN target was combined with one CV and one CVN prime; note that in doing so, CV targets and CVN targets were preceded by two different groups of characters with CV syllables and CVN syllables, respectively. For instance, the CV target 鼻子 (/bi2.zi5/, *nose*) was combined with the CV prime 彼 (/bi3/, *that*) and the CVN prime 并 (/bing4/, *and*), respectively. Likewise, the CVN target 斑马 (/ban1.ma3/, *zebra*) was combined with the CVN prime 扮 (/ban4/, *play*) and the CV prime 罢 (/ba4/, *stop*), respectively. Finally, for each CV and CVN target, an unrelated condition was formed by combining it with a phonologically unrelated prime character (e.g., 找(/zhao3/, *find*)—鼻子 (/bi2.zi5/); 内(/nei4/, *inside*)—斑马(/ban1.ma3/). Primes and the initial characters of target names always had different tones. Semantic or orthographic overlap between primes and targets was avoided. All stimuli are presented in Appendix A (as a supplementary file).

2.3 Design

The experiment adopted a 2 (Target Type: CV, CVN) \times 3 (Prime Type: CV, CVN, unrelated) within-participants design. Because of limitations on available items, we manipulated the repetition exactly in the same way in J.-Y. Chen et al. (2003). Each participant named the 40 targets three times preceded by three kinds of primes, i.e., 120 trials in one repetition, three repetitions resulting in 360 trials in total. Each repetition was set in one block, resulting in three blocks in total. The order of target pictures within a block was pseudo-randomized to prevent targets from repeating within five trials. A new sequence was generated for each participant and each block.

2.4 Apparatus

The experiment was performed using E-Prime Professional Software (Version 1.1; Psychology Software Tools). Participants were seated in a quiet room approximately 70 cm from a 21 inch CRT computer screen with a refresh rate of 100 Hz. Naming latencies were measured from target onset using a voice-key, connected to the computer via a PST Serial Response Box.

2.5 Procedure

Before the experiment, participants were asked to familiarize themselves with the target pictures by viewing each target for 2000 ms with the picture name printed below each picture. After the learning phase, participants received a picture naming test without concurrently presented names. When all pictures were named correctly, the experimental blocks were administered, comprising 120 trials per block.

Each trial involved the following sequence: a fixation cross (+) was presented at the center of the screen for 500 ms, followed by a forward mask (@@) for 500 ms. Then, the prime was presented for 50 ms, followed by a backward mask (@@) for 20ms. After that, the target picture appeared, and

it disappeared after 2 seconds or when participants made a vocal response. Primes were presented in 28-point bold Song font, and forward and backward masks in 36-point Song font. The visual angles of the targets were less than 2 degrees horizontally and vertically. Participants were asked to name the pictures aloud as quickly and accurately as possible. Following each response, the experimenter judged and recorded whether the response was correct or not (or whether a voice key error had occurred). An inter-trial interval of 1000 ms concluded each trial. The experiment took about 30 minutes in total.

2.6 EEG recordings and analysis

The electroencephalogram (EEG) was recorded with 64 electrodes secured in an elastic cap (Electro Cap International) using *Neuroscan 4.3* software. The vertical electro-oculogram (VEOG) was monitored with two electrodes placed above and below the left eye. The horizontal EOG (HEOG) was recorded by a bipolar montage using two electrodes placed on the right and left external cantus. The left mastoid electrode served as reference. All electrode impedances were kept below 5 k Ω during the experiment. Electrophysiological signals were amplified with a band-pass filter of 0.05 and 70 Hz and digitized continuously at a rate of 500 Hz.

The package *Neuroscan 4.3* was used in the ERP data analysis. The EEG data were re-referenced off-line to the average of both mastoids and filtered off-line using a 0.03 Hz high-pass filter and 40 Hz low-pass filter. The EEG data were segmented from 100 ms before to 600 ms after the onset of the pictures, with baseline correction from -100 to 0 ms preceding pictures onset. Epochs containing artifact signals below/above $\pm 100\mu\text{V}$ were rejected.

3 Results

3.1 Behavioral data

One participant was excluded because he made more than 10% errors. Incorrect responses (0.89%), voice key errors (1.60%), naming latencies longer than 1500 ms or shorter than 350 ms (0.37%), and those deviating by more than two standard deviations from each repetition and each participant's mean (5.37%) were excluded. The remaining data were used in subsequent statistical analyses. Table 2 shows mean latencies and error percentages, presented by Target and Prime type.

(Insert Table 2 here)

The behavioral data was analyzed using a linear mixed effects model (LMM) (Bates, 2005; Baayen et al., 2008), according to Barr et al.'s (2013) guideline for the maximal random effects structure. The *lmer()* function of the lme4 package was used to estimate fixed effects and parameter estimation of the LMM. The degree of freedom and p values were computed using the *anova()* function of the *lmerTest* package with Satterthwaite approximations (Kuznetsova, Brockhoff, & Christensen, 2014). These analyses were conducted using the free software R (R Development Core Team, 2013).

Syllable priming effect. The data were analyzed using a LMM that included fixed effects of prime type (CV vs. CVN), target type (CV vs. CVN), with by-participant and by-item random intercept and slope adjustments for all fixed effects. Note that the condition with unrelated primes was not included in the analysis. For latencies, the effect of prime type was not significant, $F(1, 5882.3) = 0.26$, $p = 0.61$, the effect of target type was not significant, $F(1, 37.7) = 0.25$, $p = .62$. Critically, the interaction between prime type and target type was significant, $F(1, 5882.3) = 10.10$, $p = 0.001$: as evident in Table 2, response latencies were faster when target and prime type matched, compared to

when they mismatched, which is the signature pattern of syllabic priming showed in previous studies.³

Error rates (overall less than 1%) were considered too low to allow for a meaningful statistical analysis.

Phoneme overlap effect. The data were analyzed using a LMM that included the fixed effect of phonemic overlap (phonemically related vs. unrelated) with by-participant and by-item random intercepts and slope adjustments for the fixed effect. The phonemically related condition consists of CV prime paired with CVN target and CVN prime paired with CV target, while the unrelated condition consists of two unrelated primes paired with CV target and CVN target conditions, respectively. For latencies, the effect of phonemic overlap was significant: latencies were longer in the phonemically related condition (694 ms) than in the unrelated condition (688 ms), $\beta = -4.79$, $t(5877) = -2.01$, $p = .04$.⁴ Note that phonemic overlap hence resulted in a slightly *inhibitory* effect on latencies.

3.2 ERP Data analyses

Prior to off-line averaging, all single-trial waveforms were screened for eye movements, electrode drifting, amplifier blocking and EMG artifacts. Trials in which speakers produced incorrect

³ For the latencies, in an additional analysis repetition was included as a fixed factor and was significant, $F(2, 5878.1) = 40.00$, $p < 0.001$. However, the interaction between repetition and prime type, $F(2, 5877.8) = 0.49$, $p = 0.61$, the interaction between repetition and target type, $F(2, 5878.1) = 1.81$, $p = 0.16$, and the triple interaction among prime type, target type and repetition, $F(2, 5877.8) = 0.18$, $p = 0.83$, were not significant. We therefore collapsed across the three repetitions in the behavioral data analysis reported here.

⁴ In an additional analysis, repetition was included as a fixed factor and was significant, $F(2, 5874.9) = 34.21$, $p < 0.001$, however, the interaction between repetition and phoneme overlap type, $F(2, 5864.6) = 0.80$, $p = 0.45$, the interaction between repetition and target type, $F(2, 5864.8) = 1.84$, $p = 0.15$, and the triple interaction among prime type, target type and repetition, $F(2, 5864.6) = 0.68$, $p = 0.50$, were not significant. We therefore collapsed across the three repetitions in the behavioral data analysis.

responses and those with onset latencies faster than 600 ms (20.8%) were excluded from the EEG analysis. The latter criterion was applied to avoid contamination from potential artefacts associated with articulation (see also Qu, Zhang, & Damian, 2016; Zhu, Damian & Zhang, 2015 for a similar procedure); note that this results in largely overlapping (~80%) but not identical data for the behavioral and the ERP analysis. ERPs were calculated by averaging the EEG time-locked to a point 100 ms before picture onset and lasting until 600 ms after picture onset. Mean amplitudes were calculated separately for each participant and each condition in six time windows (0-100, 100-200, 200-300, 300-400, 400-500, and 500-600 ms). Statistical analysis was carried out in six regions of interest (ROIs), with each representing an average of three electrodes: left-anterior (F3, FC3, C3), midline anterior (Fz, FCz, Cz), right anterior (F4, FC4, C4), left posterior (CP3, P3, PO3), midline posterior (CPz, Pz, POz), and right posterior (CP4, P4, PO4).

In analyzing the EEG results, we pursued the following approach. First, in parallel to the analyses conducted on the behavioral results, an omnibus ANOVA was conducted which included the two target types (CV and CVN) and two prime types (CV and CVN), and additionally region (anterior/posterior) and laterality (left, middle, right) of electrodes. The ANOVA analysis were conducted separately for six relevant post-onset time windows of 100 ms duration each, and allowed us to explore the specific interplay of prime type, target type, and location, as well as its time course. Second, we focused on the interaction between prime type (CV and CVN) and target type (CV and CVN) to investigate the temporal course of syllable priming effect, and we analyzed EEG including region and laterality, and for each relevant time window separately. Third, in order to investigate the temporal course of phoneme overlap effects, we focused on the effect of prime type as well as interactions in which prime type was involved, in an ANOVA analysis with prime type (phonemically

related: CV prime paired with CVN target, CVN prime paired with CV target; and phonemically unrelated: two unrelated primes paired with the CV and CVN targets) and target type (CV and CVN) as factors. We used the false discovery rate (*FDR*) method in the multiple comparisons (Yekutieli & Benjamini, 1999; see also Fan, Batmanghelich, Clark, Davatziko, & Alzheimer's Disease Neuroimaging Initiative, 2008), as implemented in the R software using package *fdrtool*.

3.2.1 Omnibus ANOVA: Interaction between prime and target type

Figure 2 shows grand average ERPs for CV and CVN targets (upper and lower panel respectively) when preceded by CV and CVN primes, for individual ROIs. For each of the six 100 ms time windows, mean amplitudes were entered into a 2 X 2 X 2 X 3 repeated measures ANOVA with the factors target type (CV and CVN), prime type (CV and CVN), region (anterior and posterior), and laterality (left, middle, and right). Results are shown in Table 3 (upper panel). The Greenhouse-Geisser correction was applied to all repeated measures with more than one degree of freedom. As can be seen in the Table, prime type and target type interacted with each other, or were involved in a higher-order interaction, in all time windows from 300-400 ms to 500-600 ms.

(Insert Figure 2 here)

(Insert Table 3 here)

3.2.2 Syllable priming effect

Table 3 suggests that syllabic priming (significant interactions among prime type and target type, and/or higher-order interactions involving prime and target type) were found in the time windows from 300 to 600 ms. The interaction between prime and target type was further analyzed in different time windows and different ROIs. The FDR correction was applied to the statistics in different time

windows. Table 3 (lower panel) reports the presence or absence of an interaction between prime and target type on mean amplitudes in the time windows of 300-600 ms after picture onset. A significant effect was found in the time window of 300-400 ms, mainly at right anterior, middle and right posterior regions. Figure 3 shows the interactions between prime type and target type at individual ROIs in this time window, reflecting significant syllable priming effect at right anterior and posterior, as well as middle posterior regions (all $ps < .05$ after FDR correction).

(Insert Figure 3 here)

3.2.3 Phonemic overlap effect

To obtain estimates of the time course of phonemically based priming in this task, we collapsed CV primes-CVN targets and CVN primes-CV targets conditions to form a “phonemically related” condition, and we combined trials with unrelated primes and targets to form the “unrelated” condition. Figure 4 shows grand average ERPs for the phonemically related and unrelated conditions at individual ROIs (all $ps < .05$ after FDR correction). Mean amplitudes for the phonemically related and the unrelated condition were statistically compared in six time windows (see Table 4). The FDR correction was applied in the multiple comparisons. Comparisons were significant at right posterior regions, and marginally significant at middle anterior and posterior regions, and were restricted to the time window of 500-600 ms post-stimulus.

(Insert Table 4 here)

(Insert Figure 4 here)

4 Discussion

The present study investigated syllabic and phonemic encoding in the spoken word production of Chinese participants via a masked priming paradigm. Behavioral results exhibited the

characteristic interaction between prime and target syllable type previously shown with Chinese speakers (J.-Y. Chen et al., 2003; You et al., 2012). Our finding of syllabically based priming in Chinese speech production is also broadly in line with parallel findings from other tasks in Chinese which have similarly suggested that syllables constitute important mental units (J.-Y. Chen et al., 2002; J.-Y. Chen et al., 2003; O'Seaghdha et al., 2010; You et al., 2012). As highlighted in the Introduction, the properties of Chinese greatly favor such syllabically based effects: not only does Chinese orthography represent syllables via characters, but also in spoken Mandarin Chinese syllable boundaries are unambiguous and syllables are never re-syllabified. In the ERP data, we found an interaction between prime and target syllable type on mean amplitudes which further highlights an influence of syllable structure. Critically, syllabic prime-target overlap modulated ERPs from 300 ms to 400 ms post picture onset. By contrast, phonemic overlap between prime and target showed a behavioral effect when compared to the neutral baseline which was weakly inhibitory, and phonemically based effects emerged in EEG only weakly, and in a later time window (500-600 ms post picture onset). Although the evidence with regard to phonemic overlap is admittedly less conclusive than on syllabic overlap, the overall pattern might suggest that syllables are retrieved earlier than phonemes in Chinese spoken production, an inference which would be broadly in line with the claim that syllables constitute "proximate units" (O'Seaghdha et al., 2010) for Chinese speakers, whereas phonemes are of subordinate importance.

The time interval in which we found syllable priming (300-400 ms) is roughly in agreement with the estimates introduced by Indefrey and Levelt (2004) for the time course of phonological encoding (275-455 ms). Mean latencies in our study were about 700 ms (see Table 2), which is somewhat (but not much) slower than the average naming time of 600 ms used for the estimates provided by

Indefrey and Levelt. The time course of form-related priming in the current study is also broadly in line with results of studies conducted in alphabetic languages and Mandarin Chinese. For instance, Dell'Acqua, Sessa, Peressotti et al. (2010) reported a picture-word interference task in English in which they observed facilitation in the phonologically related condition (shared phonemes) in the 250-400 ms time window. Eulitz, Hauk and Cohen (2000) found divergent ERPs in a similar time frame (250-400 ms) for phonological encoding in English. Specifically, Wang et al. (2017) found a syllable priming effect in the 200-600 ms interval in Chinese spoken production using a delayed naming task, which they interpreted as implying syllables as proximate units in phonological encoding. With regard to location, significant effects of syllabic match were found, in the time window ranging from 300-400 ms, in right anterior, as well as left, middle and right posterior, regions. The way in which primes and targets were combined in our study implies that both syllable match (CV prime paired with CV target, and CVN prime paired with CVN target) and syllable mismatch (CV prime paired with CVN target and CVN prime paired with CV target) conditions included phonemic overlap, but only the former also includes matching syllabic structures between primes and targets. This implies that the effect in the "syllable match" condition could not have been due to phonemic overlap (because this was also the case in the "phoneme overlap" condition), but instead arose due to matching content at the syllabic level.

Behavioral results showed that the effect of phonemic prime-target overlap was weakly inhibitory, compared to the unrelated condition. We note that J.-Y. Chen et al. (2016), in a similar masked priming study, also found an inhibitory effect of comparable magnitude. In EEG the effect appeared at a narrow "late" time window (500-600 ms). Several interpretations of this pattern are possible. We have so far assumed that the phonemic manipulation may reflect facilitation due to

priming of phonemes which are shared between primes and targets. However, under this account we would normally predict facilitation (not inhibition), and the time window in which the effect was found is clearly out of alignment with the time course of phonological encoding (275-400 ms) estimated by Indefrey and Levelt (2004). An alternative was proposed by J.-Y. Chen et al.: phonemic inhibition in masked priming could arise because of competition between similar syllables during selection of the first syllable of the target; in other words, despite a phonemic manipulation, the effect really reflects a syllabic locus. This explanation would account for behavioral inhibition, but it would not be obvious why the phonemic effect, if arising from syllable overlap, emerges at such a late point in time. Yet another explanation is that the effect reflects phonetic encoding, i.e., the speaker's preparation of a "gestural score". Phonetic encoding in speech production involves the computation of a gestural score which specifies motor tasks such as closing the glottis or releasing lip closure. In our study, the time frame of the phonemic effect is roughly in agreement with the estimated time course of phonetic encoding (455-600 ms) suggested by Indefrey and Levelt (2004), and it has been previously suggested (Kinoshita, 2000) that form priming in masked priming tasks arises due to phonetic overlap. Finally, the phonemic overlap effect could reflect internal speech monitoring for phonological encoding. Self-monitoring of speech production involves an internal loop that monitors abstract phonological codes, and an external loop that monitors self-generated spoken languages (Levelt et al., 1999). Qu et al. (2012) found a more negative ERP amplitude in the phonologically related condition in the 300-400 ms interval after picture onset (see also Yu et al., 2014, for a similar finding), and they proposed that it might reflect internal speech monitoring. The latencies in the present study were around 690 ms, which were somewhat longer than those typically found in picture naming studies (which tend to be around 600 ms), hence the phoneme

priming effect in the time window of 500-600 ms might arise as a result of monitoring. Overall, it is acknowledged that evidence from the phonemic overlap condition in the current study is currently inconclusive regarding its underlying locus and mechanism.

Some aspects in which our findings diverge from previous EEG studies on spoken production should be highlighted. Both the syllabically and phonemically related condition were characterized by less positive ERP waveforms when compared to the unrelated condition. This pattern is consistent with a number of previous studies (e.g., Thierry & Wu, 2004; Wu & Thierry, 2010; Liotti, Woldorff, Perez III, & Mayberg, 2000; Wang et al., 2017) in which similarly less positive ERPs were found in the phonologically congruent compared to an incongruent condition. On the other hand, Dell'Acqua et al. (2010) found that in a picture-word interference study, phonologically related distractors induced a *more* positive going waveform than unrelated distractors. In a study of spoken adjective-noun production, Qu et al. (2012) found a complex pattern, namely a more positive going waveform in the time window of 200-300 ms, but a less positive going waveform in the time window of 300-400 ms (see also Yu et al., 2014), and they speculated that the early pattern might reflect phonological encoding proper, whereas the later difference reflects internal self-monitoring. The waveforms directions and the map distribution of the syllable priming effect reported here were quite similar to those observed by Wang et al. (2017), and a picture naming task was used in both studies. Different paradigms (picture naming in Wang et al.'s and the current study, but phrase production in Qu et al. and Yu et al.) probably elicit different map distributions. O'Seaghdha et al. (2013) also pointed out that the observed effects in phrase production experiments might not necessarily reflect the functional engagement of phonemes in spoken word production. Further research is needed to elucidate these patterns, particularly given the different utterance formats as well as response

languages.

The current study adds to a growing body of evidence showing important differences between spoken production in Indo-European languages, and Chinese. In the Introduction, we noted that the particular prominence of the syllable, and by comparison the subsidiary role of the phoneme, for Chinese speakers might arise due to various reasons. First, the specific properties of spoken Mandarin (i.e., relatively few syllable types, clear syllabic boundaries) render syllables particularly salient; second, Chinese orthography primarily represents syllables (whereas alphabetic orthographies represent speech sounds). Of course, these two scenarios could also be causally related, e.g., Chinese orthography primarily represents syllables because spoken Chinese has clear syllabic structure. Our results reported here do not offer directly relevant evidence, and we believe extensive cross-linguistic comparison between languages of different types and with different orthographic systems would be required to resolve this issue.

To conclude, the current behavioral and ERP results provide consistent evidence for the claim that syllables constitute proximate functional units of spoken production for Mandarin speakers. The fine-grained time course of syllable and phoneme overlap effects afforded by EEG provides important steps toward a temporal map of speech production.

Acknowledgments

This research was supported by grants from the National Natural Science Foundation of China (31471074), the Key project of Beijing Social Science Foundation in China (16YYA006), and the Fundamental Research funds for the Central Universities; and the Research Funds of Renmin University of China (18XNLG28) to Qingfang Zhang.

References

- Alario, F.-X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition*, 102, 464–475. <https://doi.org/10.1016/j.cognition.2006.02.002>
- Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. (2005). Fitting linear mixed models in R. *R News*, 5(1), Article 5. Available: <http://modeloslineales.wdfiles.com/local--files/start/bates.2005.pdf>
- Beijing Language Institute (1986). *Modern Chinese frequency dictionary (in Chinese)*. Beijing, China: Beijing Language Institute Publisher.
- Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, 123, 84–99. <https://doi.org/10.1016/j.cognition.2011.12.007>
- Brand, M., Rey, A., & Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory and Language*, 48, 435–443. [https://doi.org/10.1016/S0749-596X\(02\)00522-3](https://doi.org/10.1016/S0749-596X(02)00522-3)
- Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, 90, 393–340. <https://doi.org/10.1016/j.bandl.2003.12.003>
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive*

Neuropsychology, 14, 177–208. <https://doi.org/10.1080/026432997381664>

Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese.

Psychologia: An International Journal of Psychology in the Orient, 43, 15–26.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46(4), 751–781.

<http://dx.doi.org/10.1006/jmla.2001.2825>

Chen, J. Y., O'Seaghdha, P. G. & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 42, 825–836.

<http://dx.doi.org/10.1037/a0039911>

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology*, 45(1), 107–120.

<http://dx.doi.org/10.6129/CJP.2003.4501.07>

Chetail, F., & Mathey, S. (2009). Syllabic priming in lexical decision and naming tasks: the syllable congruency effect re-examined in French. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale*, 63(1), 40–48. <http://dx.doi.org/10.1037/a0012944>

Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表 [Chinese Lexicon]

(CLDC-LAC-2003-01). Beijing, China: State Key Laboratory of Intelligent Technology and Systems, Tsinghua University, and Institute of Automation, Chinese Academy of Sciences.

Cholin, J., Levelt, W., & Schiller, N. (2006). Effects of syllable frequency in speech production.

Cognition, 99(2), 205–235. <https://doi.org/10.1016/j.cognition.2005.01.009>

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word

production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning*

- Memory and Cognition*, 37, 109–122. <http://dx.doi.org/10.1037/a0021322>
- Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, 1147, 192–208. <https://doi.org/10.1016/j.brainres.2007.01.137>
- Costa, A., & Sebastian-Galles, N. (1998). Abstract phonological structure in language production: Evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 886. <http://dx.doi.org/10.1037/0278-7393.24.4.886>
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody and the comprehension of spoken language: A literature review. *Language and Speech*, 40(2), 141–201. <https://doi.org/10.1177/002383099704000203>
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385–400. [https://doi.org/10.1016/0749-596X\(86\)90033-1](https://doi.org/10.1016/0749-596X(86)90033-1)
- Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, 49, 119–132. [https://doi.org/10.1016/S0749-596X\(03\)00008-1](https://doi.org/10.1016/S0749-596X(03)00008-1)
- Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, 57, 195–209. <https://doi.org/10.1016/j.jml.2006.11.001>
- Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, 24, 685–712. <https://doi.org/10.1080/01690960802351260>
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. <http://dx.doi.org/10.1037/0033-295X.93.3.283>

- Dell, G. S. (1988). The retrieval of phonological forms in production: tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124–142.
[https://doi.org/10.1016/0749-596X\(88\)90070-8](https://doi.org/10.1016/0749-596X(88)90070-8)
- Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Frontiers in Psychology*, 1, 177. <https://doi.org/10.3389/fpsyg.2010.00177>
- Euilz, C., Hauk, O., & Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clinical Neurophysiology*, 111, 2088–2977.
[https://doi.org/10.1016/S1388-2457\(00\)00441-7](https://doi.org/10.1016/S1388-2457(00)00441-7)
- Fan, Y., Batmanghelich, N., Clark, C. M., Davatzikos, C., & Alzheimer's Disease Neuroimaging Initiative. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage*, 39(4), 1731–1743.
<https://doi.org/10.1016/j.neuroimage.2007.10.031>
- Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: the role of syllabic units. *Journal of Memory and Language*, 35, 708–723.
<https://doi.org/10.1006/jmla.1996.0037>
- Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory and Cognition*, 25, 458–470. <https://doi.org/10.3758/BF03201122>
- Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, 5, 67–103. <https://doi.org/10.1080/02643298808252927>
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>

- Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin & Review*, 7(1), 133–141. <https://doi.org/10.3758/BF03210732>
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models. R package, version 2.0-3.
- Laganaro, M., Morand, S., & Schnider, A. (2009). Time course of evoked-potential changes in different forms of anomia in aphasia. *Journal of Cognitive Neuroscience*, 21(8), 1499–1510. <https://doi.org/10.1162/jocn.2009.21117>
- Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13464–13471. <https://doi.org/10.1073/pnas.231459498>
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269. [https://doi.org/10.1016/0010-0277\(94\)90030-2](https://doi.org/10.1016/0010-0277(94)90030-2)
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75. <http://dx.doi.org/10.1017/S0140525X99001776>
- Liotti, M., Woldorff, M. g., Perez III, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, 38, 701–711. [http://dx.doi.org/10.1016/S0028-3932\(99\)00106-2](http://dx.doi.org/10.1016/S0028-3932(99)00106-2)
- Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 19558–19563. <https://doi.org/10.1073/pnas.0607065104>

- Maess, B., Friederici, A. D., Damian, M. F., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, 14, 455–462. <http://dx.doi.org/10.1162/089892902317361967>
- Mehler, J., Dommergues, T. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal behavior*, 20, 298–305.
[https://doi.org/10.1016/S0022-5371\(81\)90450-3](https://doi.org/10.1016/S0022-5371(81)90450-3)
- Meng, X., Sai, X., Wang, C., Wang, J., Sha, S., & Zhou, X. (2005). Auditory and speech processing and reading development in Chinese school children: behavioral and ERP evidence. *Dyslexia*, 11, 292–310. <https://doi.org/10.1002/dys.309>
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language* 30, 69–89.
[http://dx.doi.org/10.1016/0749-596X\(91\)90011-8](http://dx.doi.org/10.1016/0749-596X(91)90011-8)
- O'Séaghdha, P. G. (2015). Across the great divide: Proximate units at the lexical-phonological interface. *Japanese Psychological Research*, 57, 4–21. <http://dx.doi.org/10.1111/jpr.12074>
- O'Séaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, 115, 282–302. <https://doi.org/10.1016/j.cognition.2010.01.001>
- O'Séaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2013). Close but not proximate: The significant of phonological segments in speaking depends on their functional engagement. *Proceedings of the National Academy of Sciences of the United States of America*, 110, E3.
<https://doi.org/10.1073/pnas.1217032110>
- Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin

- speakers. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 14265–14270. <https://doi.org/10.1073/pnas.1200632109>
- Qu, Q., Zhang, Q., & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language*, 159, 118–126. <https://doi.org/10.1016/j.bandl.2016.06.008>
- R Development Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Roelofs, A. (1997a). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, 12, 657–694. <https://doi.org/10.1080/016909697386655>
- Roelofs, A. (1997b). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284. [http://dx.doi.org/10.1016/S0010-0277\(97\)00027-9](http://dx.doi.org/10.1016/S0010-0277(97)00027-9)
- Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research*, 57, 22–37. <http://dx.doi.org/10.1111/jpr.12050>
- Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, 39, 484–507. <https://doi.org/10.1006/jmla.1998.2577>
- Schiller, N. O. (1999). Masked syllable priming of English nouns. *Brain and Language*, 68, 300–305. <https://doi.org/10.1006/brln.1999.2109>
- Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 512–528. <http://dx.doi.org/10.1037/0278-7393.26.2.512>

- Schiller, N. O., Costa, A., & Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. *Papers in Laboratory Phonology*, 7, 35–59.
- Sevald, C. A., Dell, G. S., & Cole, J. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, 34, 807–820.
<http://dx.doi.org/10.1006/jmla.1995.1035>
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Erlbaum.
- Thierry, G., & Wu, Y. (2004). Electrophysiological evidence for language interference in late bilinguals. *Neuroreport*, 15, 1555–1558. <http://dx.doi.org/10.1097/01.wnr.0000134214.57469.c2>
- Verdonschot, R.G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W., & Schiller, N.O. (2011). The functional unit of Japanese word naming: evidence from masked priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37, 1458–1473.
<http://dx.doi.org/10.1037/a0024491>
- Verdonschot, R. G., Lai, J., Feng, C., Tamaoka, K., & Schiller, N. O. (2015). Constructing initial phonology in Mandarin Chinese: syllabic or sub-syllabic? A masked priming investigation. *Japanese Psychological Research*, 57, 61–68. <http://dx.doi.org/10.1111/jpr.12064>
- Wong, A. W.-K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1172–1190. <http://dx.doi.org/10.1037/a0013000>
- Wong, A. W.-K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review*, 16, 888–892.

<http://dx.doi.org/10.3758/PBR.16.5.888>

- Wang, J., Wong, A. W.-K., Wang, S., & Chen, H.-C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports*, 7, 5815. <http://dx.doi.org/10.1038/s41598-017-06186-z>
- Wu, Y., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *The Journal of Neuroscience*, 30(22), 7646–7651. <http://dx.doi.org/10.1523/JNEUROSCI.1602-10.2010>
- You, W., Zhang, Q., & Verdonchot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS ONE*, 7(10): e46595. <http://dx.doi.org/10.1371/journal.pone.0046595>
- Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82, 171–196. [https://doi.org/10.1016/S0378-3758\(99\)00041-5](https://doi.org/10.1016/S0378-3758(99)00041-5)
- Yu, M., Mo, C., & Mo, L. (2014). The role of phoneme in Mandarin Chinese Production: Evidence from ERPs. *PLoS ONE*, 9(9): e106486. <http://dx.doi.org/10.1371/journal.pone.0106486>
- Yue, Y. & Zhang, Q. (2015). Syllable and segments effects in Mandarin Chinese spoken word production (in Chinese). *Acta Psychologica Sinica*, 47(3), 319-328. DOI: 10.3724/SP.J.1041.2015.00319
- Zhu, X., Damian, F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language*, 144, 16–25. <http://dx.doi.org/10.1016/j.bandl.2015.03.007>

Table 1. Property of the stimuli used in Experiment

	NA	WF	(I)CF	(I)SF	(I)SN
Target names					
CV target	72.6	4.27	590	1093	8
CVN target	72.8	4.25	615	1112	10
Primes paired with CV target					
CV prime		--	196	838	8.55
CVN prime		--	194	834	9.30
Unrelated		--	196	844	8.55
Primes paired with CVN target					
CV prime		--	198	844	9.00
CVN prime		--	195	837	8.55
Unrelated		--	194	844	9.15

Note: NA = name agreement (in %); WF = word frequency (per million); ICF = initial character frequency (per million) of target; ISF = initial syllable frequency (per million) of target; ISN = stroke number of initial character of target; CF = character frequency (per million); SN = stroke number; SF = syllable frequency (per million).

Table 2. Mean naming latencies (in ms) and errors (in percent) with standard deviation in parentheses.

Primes	Latencies (SD)		Error rate (%) (SD)	
	CV target	CVN target	CV target	CVN target
CV	688 (61)	692 (57)	0.75 (1.17)	0.74 (1.33)
CVN	696 (63)	683 (64)	0.94 (1.63)	1.05 (1.68)
Unrelated	692 (62)	684 (64)	0.63 (0.95)	1.23 (1.57)

Table 3. Omnibus ANOVA with the factors target type (CV and CVN), prime type (CV and CVN), region (anterior and posterior), and laterality (left, middle, and right) and interactions between prime and target type at individual ROIs, in different time windows (in ms). Values indicate F statistics.

Source	Time Windows (in ms)					
	0-100	100-200	200-300	300-400	400-500	500-600
Target (1, 26)	—	13.74***	—	—	9.14**	7.02*
Prime (1, 26)	—	—	—	—	—	—
Target x Prime (1,26)	—	—	—	4.45*	3.32†	3.25†
Target x Region (1, 26)	—	—	5.09*	—	—	—
Prime x Region (1, 26)	—	—	4.39*	8.70**	4.43*	—
Target x Laterality (2, 52)	—	—	3.72*	8.54**	—	—
Prime x Laterality (2, 52)	—	—	—	—	—	—
Target x Prime x Laterality (2, 52)	—	—	—	3.58*	2.67†	—

Target x Prime x Region (2, 52)	—	—	—	—	—	—
Target x Region x Laterality (2, 52)	—	—	—	—	3.48*	—
Prime x Region x Laterality (2, 52)	—	—	—	—	—	—
Target x Prime x Region x Laterality (2, 52)	—	—	—	3.26*	2.74†	4.08*

The interaction between prime and target at individual ROIs

Left anterior (1, 26)	—	—	—			
Midline anterior (1, 26)	—	—	—			
Right anterior (1, 26)	—	—	—	5.10*	—	—
Left posterior (1, 26)	—	—	—	3.68†	—	—
Midline posterior (1, 26)	—	—	—	6.53*	—	—
Right posterior (1, 26)	—	—	—	6.15*	—	—

Note: ** $p < 0.01$, * $p < 0.05$, † $0.05 < p < 0.1$ (after the *fdr* correction)

Table 4. Comparison of phonemically related and unrelated conditions at individual ROIs, in different time windows (in ms). Values indicate *F* statistics.

	Time Windows (in ms)					
	0-100	100-200	200-300	300-400	400-500	500-600
Left anterior (1,26)	—	—	—	—	—	—
Midline anterior (1,26)	—	—	—	—	—	2.29 [†]
Right anterior (1,26)	—	—	—	—	—	—
Left posterior (1,26)	—	—	—	—	—	—
Midline posterior (1,26)	—	—	—	—	—	2.01 [†]
Right posterior (1,26)	—	—	—	—	—	2.56 [*]

Note: ^{*} $p < 0.05$, [†] $0.05 < p < 0.1$ (after the *fdr* correction).

Figure Captions

Figure 1. Word-form encoding of spoken word production in Mandarin Chinese (adopted from O'Seaghdha, J.-Y. Chen, & T.-M. Chen, 2010).

Figure 2. Grand average ERPs for CV and CVN targets (upper and lower panel respectively) when preceded by CV and CVN primes, for individual ROIs.

Figure 3. Interaction between prime type and target type in the time window of 300-400 ms at individual ROIs ($^{\dagger}p < .10$, $*p < .05$ after FDR correction).

Figure 4. Grand average ERPs for phonemically related and unrelated primes and targets, for individual ROIs.

Figure 1.

Concept



Word

Syllable

Segments

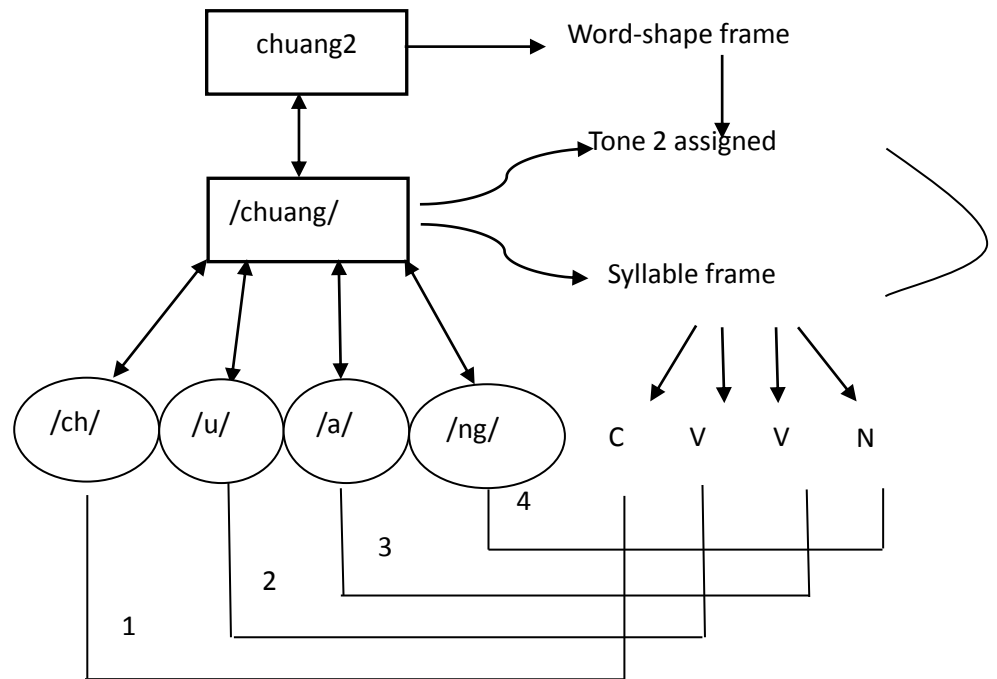


Figure 2.

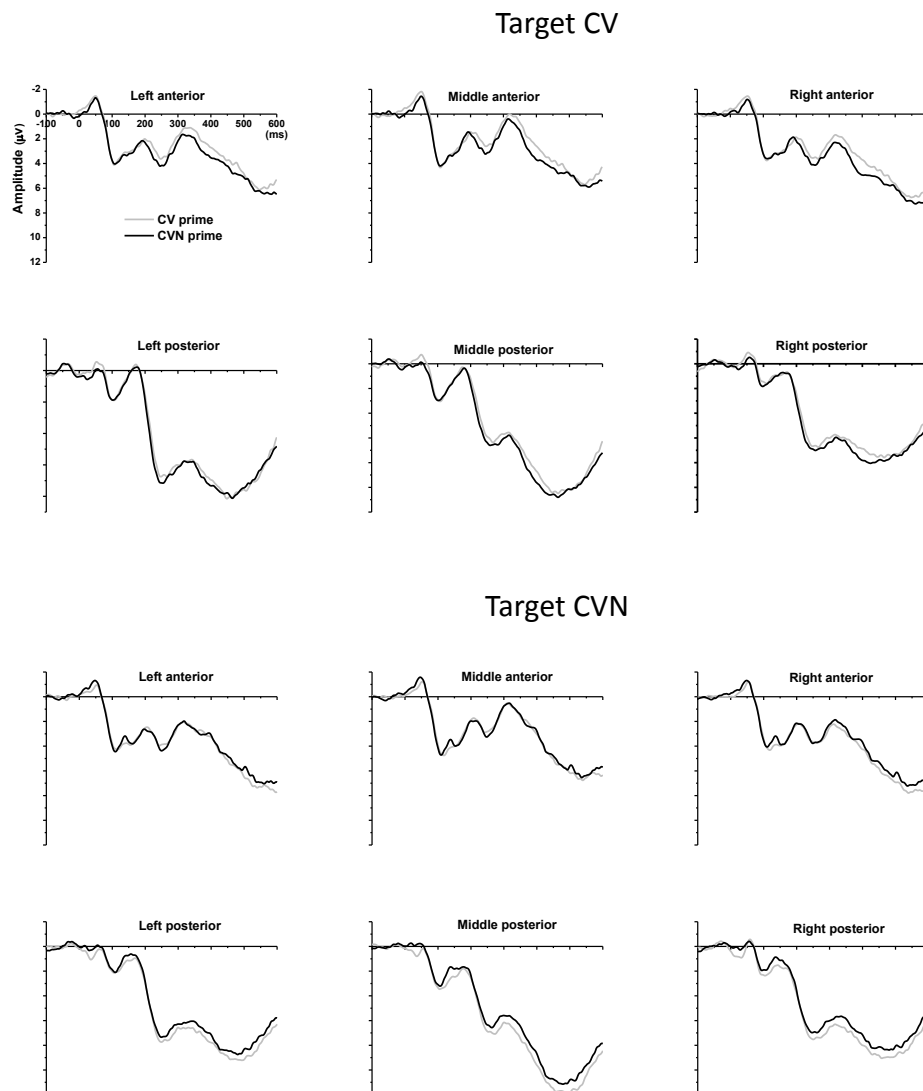


Figure 3.

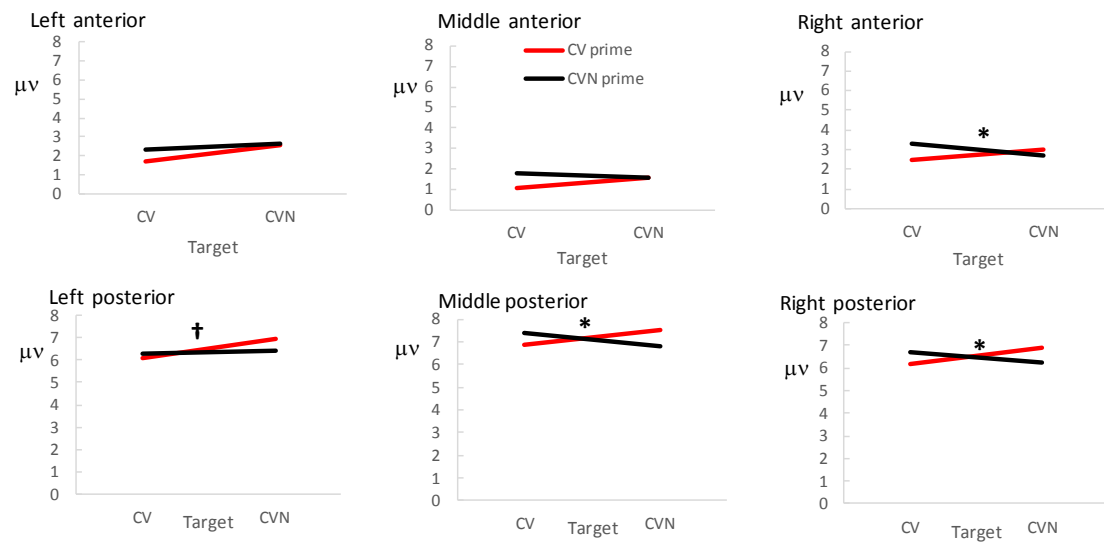


Figure 4.

